

Normalized Cone Regression

Jan de Leeuw

First created on March 28, 2019. Last update on April 29, 2019

Abstract

We discuss projection on the intersection of a polyhedral convex cone and a sphere, in particular when the target is in the polar cone. We also discuss projection on the double cone formed by the cone and its negative.

Contents

1	Introduction	1
2	History	2
3	Unnormalized Cone Projection	3
4	Normalized Cone Regression	4
4.1	Target not in Polar Cone	4
4.2	Target in Polar Cone	5
5	Double Cone Regression	6
	References	7

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory deleeuwpx.net/pubfolders/dcone has a pdf version, the bib file, and the complete Rmd file.

1 Introduction

Suppose \mathcal{E} is a finite-dimensional Euclidean space with inner product $\langle \bullet, \bullet \rangle$ and corresponding norm $\| \bullet \|$. Also \mathcal{S} is the unit sphere in \mathcal{E} , and \mathcal{B} is the unit ball.

The Gifi framework for descriptive multivariate analysis (Gifi (1990), Michailidis and De Leeuw (1998), De Leeuw and Mair (2009)) covers both linear and bilinear nonmetric multivariate methods. In linear nonmetric techniques such as nonmetric regression, discriminant analysis, and additivity analysis we have to solve the problem of minimizing $\sigma(y, x) = \|x - y\|^2$ over $y \in \mathcal{L}$ and over $x \in \mathcal{K} \cap \mathcal{S}$, where $\mathcal{K} \subseteq \mathcal{E}$ is a closed and pointed polyhedral convex cone and $\mathcal{L} \subseteq \mathcal{E}$ is a linear subspace. In bilinear nonmetric methods such as principal component and canonical analysis this minimization problems has to be solved many times.

The standard way to solve this minimization problem is to use *alternating least squares*. We alternate finding the optimal y for given x and the optimal x for given y until convergence. The optimal y for given x is just a simple linear least squares problem. The optimal x for given y is a *normalized cone regression* problem (De Leeuw (1975)). It can *usually* be solved by first projecting y on the cone \mathcal{K} and then normalizing the projection to unit length.

It has typically not been emphasized in the data analysis literature that normalized cone regression can go astray, especially in early iterations and in situations where the least squares fit is very bad. In these, admittedly rare, cases the projection of y on the cone \mathcal{K} is the zero vector, and we cannot normalize a zero vector.

2 History

Gifi (1990) briefly discusses the problem on pages 529-530, which happen to be the very last pages of the book, right before the references and the index. We give the relevant quotation, for those who do not want to spend the \$ 330.95 to buy the book.

In the book we sometimes use normalized cone regression. This can be either one of two things. In the first place minimization of $(y - x)'W(y - x)/x'Wx$ over x in a cone C . This is called *implicit normalization*. This name suggests there is also something like *explicit normalization*. This is minimization of $(y - x)'W(y - x)$ over all $x \in C$ that satisfy in addition $x'Wx=1$. It is a basic result of Kruskal and Carroll (1969) that in this simple case implicit normalization, explicit normalization, and no normalization all give essentially the same solution. All solutions are proportional to the projection of y on the cone, with only the proportionality constant different for the different problems. The result does not rely on convexity. There is one exception which should be noted. If y is in C^\ominus , i.e. $y'Wz \leq 0$ for all $z \in C$, then the origin is the projection of y on C . In the normalized problems the solution of x is one of the extreme rays of C , suitably normalized. An extreme ray is any ray in the cone that cannot be written as a nonnegative linear combination of two other rays in the cones. If C is a subspace and y is in the W -orthogonal complement, then the infimum in the implicitly normalized problem is one - not attained, but approached by letting $x \rightarrow \infty$. The infimum in the explicitly normalized problem is attained for any $x \in C$ with $x'Wx = 1$; it is equal to $1 + y'Wy$. If C is the cone used in monotone regression, then the extreme rays are the vectors with the first k elements equal to zero and the last $n - k$ elements equal to one ($k = 1, \dots, n - 1$) together with the vectors u and $-u$ which span the intersection of C and C^\ominus . Thus the cone is not pointed. The normalized regression problem must be solved by testing all these rays and by keeping the best one.

Example 5.5.2 in Lange (2016) discusses the special case in which K is the non-negative orthant. On page 142 we see

The constraint set C in question is the intersection of the unit sphere and the nonnegative orthant. Projection of an external point y onto C splits into several

cases. When all components of y are negative, $P_C(y) = e_i$, where y_i is the least negative component of y , and e_i is the standard unit vector along coordinate direction i . The origin \mathcal{O} is equidistant from all points of C . If any component of y is positive then the projection is constructed by setting the negative components of y equal to 0 and standardizing the truncated version of y to have norm 1.

Neither Gifi nor Lange provide an actual proof for their statements. Exercise 5.6.14 in Lange (2016) asks the reader to come up with a proof. All the machinery and proofs one can possibly want were recently made available in Bauschke, Bui, and Wang (2018). That article provides a very thorough and complete analysis of projection on the intersection of a cone and a sphere (or a ball). On page 2158 the authors say

Inspired by an example in the recent and charming book [12], *our aim in this paper is to systematically study the case in which K is a closed convex cone and S is either the (convex) unit ball or (nonconvex) unit sphere centered at the origin.*

The “recent and charming” book in the quotation is, of course, Lange (2016). The authors do not refer to Gifi, which is probably because Gifi’s book is neither recent nor charming.

3 Unnormalized Cone Projection

We review some basic facts about *unnormalized* cone projection problems. Any book on convex analysis (for example Rockafellar (1970)) will provide the necessary discussion and proofs.

Definition 1: [Projection] The *projection* $P_A(x)$ of x on a set $A \subseteq \mathcal{E}$ is the set of all $y \in A$ such that $\|x - y\| = \min_{z \in A} \|x - z\|$. Projections need not be singletons and can be empty.

Definition 2: [Polar Cone] The *polar cone* K^\ominus of a cone K is the set of all y such that $\langle x, y \rangle \leq 0$ for all $x \in K$.

Result 1: [Cone Projection]

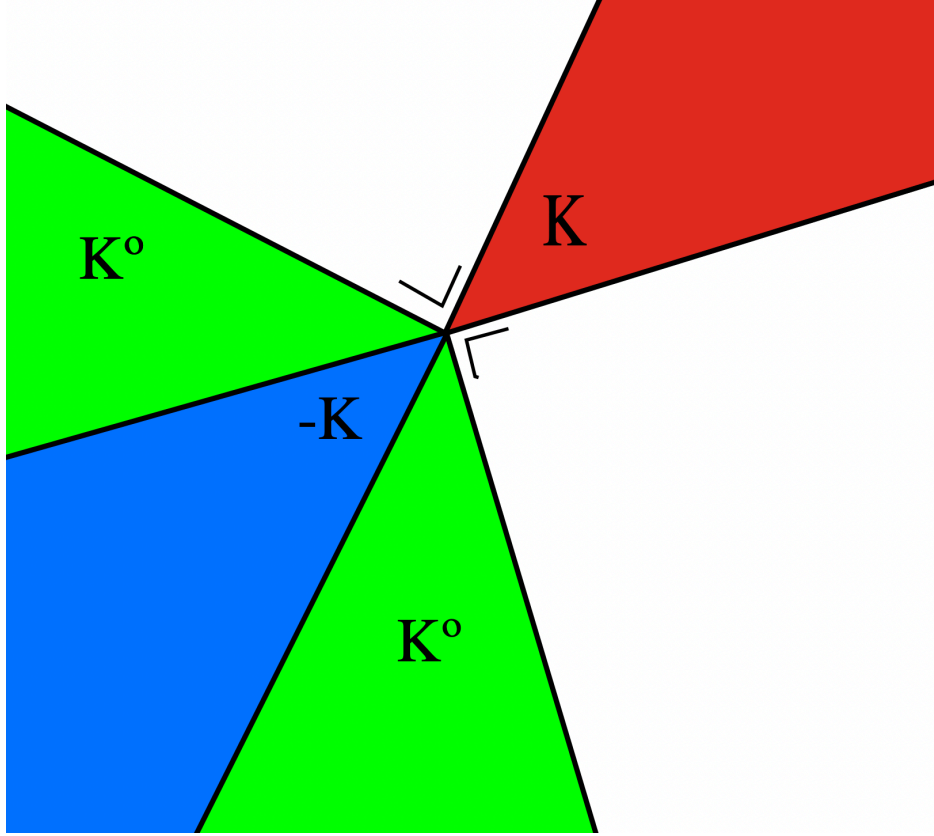
1. If K is convex the projection is a singleton. In that case we also use $P_K(x)$ for the unique element of the projection.
2. $y = P_K(x)$ if and only if $\langle y, x - y \rangle = 0$ and $\langle x, x - y \rangle \leq 0$ for all $x \in K$. Thus $y = P_K(x)$ if and only if $x - y = P_{K^\ominus}(x)$.
3. $x = P_K(x) + P_{K^\ominus}(x)$ or $P_{K^\ominus}(x) = x - P_K(x)$.
4. $\|x\|^2 = \|P_K(x)\|^2 + \|P_{K^\ominus}(x)\|^2$
5. The projection $P_K(x)$ of x on K is zero if and only if $x \in K^\ominus$.

Here are two more definitions for later use.

Definition 3: [Negative of a Cone] The *negative* K^- of a cone K is the set of all y such that $y = -x$ for some $x \in K$.

Definition 4: [Double of a Cone] The *double* K^* of a cone K is the double cone $K \cup K^-$.

In the figure below the cone is red, the polar is the union of green and blue, the negative is blue, and the double is the union of red and blue.



4 Normalized Cone Regression

We give a brief summary of the results in De Leeuw (1975) with somewhat different proofs. We discuss the relations between unnormalized, explicitly normalized, and implicitly normalized cone regression. Similar results were first discussed, somewhat informally, in Kruskal and Carroll (1969).

4.1 Target not in Polar Cone

Theorem 1: [Unnormalized]

1. If $y \in \mathcal{K}^\ominus$ then $\min_{x \in \mathcal{K}} \|x - y\|^2 = \|y\|^2$, and the minimum is attained for $x = 0$.
2. If $y \in \mathcal{E} \setminus \mathcal{K}^\ominus$ then $\mathbf{argmin}_{x \in \mathcal{K}} \|x - y\|^2 = \langle \hat{x}, y \rangle \hat{x}$, with $\hat{x} \in \mathbf{argmax}_{x \in \mathcal{K} \cap \mathcal{S}} \langle x, y \rangle$.

Proof: All we need is the following simple decomposition of the problem.

$$\begin{aligned} \min_{x \in \mathcal{K}} \|x - y\|^2 &= \min_{x \in \mathcal{K} \cap \mathcal{S}} \min_{\alpha \geq 0} \|\alpha x - y\|^2 = \\ &= \min_{x \in \mathcal{K} \cap \mathcal{S}} \min_{\alpha \geq 0} \{\alpha^2 + \|y\|^2 - 2\alpha \langle x, y \rangle\} = \|y\|^2 - \max_{x \in \mathcal{K} \cap \mathcal{S}} \max(0, \langle x, y \rangle)^2. \quad (1) \end{aligned}$$

If $y \in \mathcal{K}^\ominus$ then $\max(0, \langle x, y \rangle) = 0$ for all $x \in \mathcal{K}$. ■

Corollary 1: [Non-Polar Target]

If $y \in \mathcal{E} \setminus \mathcal{K}^\ominus$ then

$$P_{\mathcal{K} \cap \mathcal{S}}(x) = \frac{P_{\mathcal{K}}(x)}{\|P_{\mathcal{K}}(x)\|}.$$

Proof: By theorem 1 we have $P_{\mathcal{K}}(x) = \langle \hat{x}, y \rangle \hat{x}$, where $\hat{x} = P_{\mathcal{K} \cap \mathcal{S}}$. Thus $\|P_{\mathcal{K}}(x)\| = \langle \hat{x}, y \rangle$, and the result follows. ■

Theorem 2: [Explicit Normalization]

$$\operatorname{argmin}_{x \in \mathcal{K} \cap \mathcal{S}} \|x - y\|^2 = \operatorname{argmax}_{x \in \mathcal{K} \cap \mathcal{S}} \langle x, y \rangle.$$

Proof: In this case

$$\min_{x \in \mathcal{K} \cap \mathcal{S}} \|x - y\|^2 = 1 + \|y\|^2 - 2 \max_{x \in \mathcal{K} \cap \mathcal{S}} \langle x, y \rangle.$$

■

Theorem 3: [Implicit Normalization]

$$\operatorname{argmin}_{x \in \mathcal{K}} \frac{\|x - y\|^2}{\|x\|^2} = \operatorname{argmax}_{x \in \mathcal{K} \cap \mathcal{S}} \langle x, y \rangle.$$

Proof: In this case

$$\begin{aligned} \min_{x \in \mathcal{K}} \frac{\|x - y\|^2}{\|x\|^2} &= \min_{x \in \mathcal{K} \cap \mathcal{S}} \min_{\alpha \geq 0} \frac{\alpha^2 + \|y\|^2 - 2\alpha \langle x, y \rangle}{\alpha^2} = \\ &= \min_{x \in \mathcal{K} \cap \mathcal{S}} \min_{\beta \geq 0} 1 + \beta^2 \|y\|^2 - 2\beta \langle x, y \rangle = 1 - \frac{1}{\|y\|^2} \max_{x \in \mathcal{K} \cap \mathcal{S}} \max(0, \langle x, y \rangle)^2 \end{aligned}$$

■

Thus when $y \in \mathcal{E} \setminus \mathcal{K}^\ominus$, for both normalized and unnormalized cone projection, it suffices to compute

$$\operatorname{argmax}_{x \in \mathcal{K} \cap \mathcal{S}} \langle x, y \rangle = \operatorname{argmax}_{x \in \mathcal{K} \cap \mathcal{B}} \langle x, y \rangle,$$

i.e. to compute the support function of the compact convex set $\mathcal{K} \cap \mathcal{B}$ at y .

4.2 Target in Polar Cone

Theorem 4: [Polar Target] Suppose \mathcal{K} is the set of all non-negative linear combinations of m distinct unit-length vectors z_j . If $y \in \mathcal{K}^\ominus$ then

$$\max_{x \in \mathcal{K} \cap \mathcal{S}} \langle y, x \rangle = \max_{j=1}^m \langle y, z_j \rangle.$$

Proof: We have $x \in \mathcal{K} \cap \mathcal{S}$ if and only if

$$x = \alpha(\lambda) \sum_{j=1}^m \lambda_j z_j,$$

where the $0 \leq \lambda_j \leq 1$, and the λ_j add up to one, and where

$$\alpha(\lambda) = \frac{1}{\|\sum_{j=1}^m \lambda_j z_j\|}.$$

Because the norm is strictly convex we have

$$\left\| \sum_{j=1}^m \lambda_j z_j \right\| \leq \sum_{j=1}^m \lambda_j \|z_j\| = 1, \quad (2)$$

and thus $\alpha(\lambda) \geq 1$ with equality if and only if exactly one of the λ_j is equal to one (and the others are zero).

Because $y \in \mathcal{K}^\ominus$ in addition

$$\sum_{j=1}^m \lambda_j \langle y, z_j \rangle \leq \max_{j=1}^m \langle y, z_j \rangle \leq 0. \quad (3)$$

Thus, from (2) and (3),

$$\langle y, x \rangle = \alpha(\lambda) \sum_{j=1}^m \lambda_j \langle y, z_j \rangle \leq \max_{j=1}^m \langle y, z_j \rangle.$$

■

5 Double Cone Regression

If we come across an instance, in our alternating least squares iterative processes, where the target is in the polar cone it may be too much work to test all extreme rays of the cone to find the projection on $\mathcal{K} \cap \mathcal{S}$. There is, however, a convenient way around the problem.

Remember that in our algorithms we alternate minimization of $\|x - y\|^2$ over $y \in \mathcal{L}$ for fixed x and over $x \in \mathcal{K} \cap \mathcal{S}$ for fixed y . Thus we are in trouble if in some iteration we have $y \in \mathcal{L} \cap \mathcal{K}^\ominus$. The ad hoc solution, which is basically what is used in some of the Gifi programs, is to compute $P_{\mathcal{K} \cap \mathcal{S}}(-y)$ and then replace y by $-y$. This essentially means we project y on the intersection of \mathcal{S} and the double cone $K^* = \mathcal{K} \cup \mathcal{K}^-$.

Theorem 5: [Negative]

$$P_{\mathcal{K}^-}(x) = P_{\mathcal{K}}(-x)$$

Proof: We have $y = P_{\mathcal{K}}(-y)$ if and only if $\langle x, x + y \rangle \leq 0$ for all $x \in \mathcal{K}$ and $\langle y, x + y \rangle = 0$. We have $y = P_{\mathcal{K}^-}$ if and only if $\langle x, x - y \rangle \leq 0$ for all $x \in \mathcal{K}^-$, i.e. if and only if $\langle -x, -x - y \rangle = \langle x, x + y \rangle \leq 0$ for all $x \in \mathcal{K}$. ■

In an actual implementation we can follow two strategies. The first is to compute $P_{\mathcal{K}}(y)$ and $P_{\mathcal{K}^-}(y) = P_{\mathcal{K}}(-y)$ and normalize the one with the best fit. The second strategy, which is slightly more economical, is to compute $P_{\mathcal{K}}(y)$. If this is nonzero, we normalize it. If it is zero, we compute $P_{\mathcal{K}}(-y)$ and normalize it (and replace y by $-y$).

References

- Bauschke, H. H., M. N. Bui, and X. Wang. 2018. “Projecting onto the Intersection of a Cone and a Sphere.” *SIAM Journal of Optimization* 28 (3): 2158–88.
- De Leeuw, J. 1975. “A Normalized Cone Regression Approach to Alternating Least Squares Algorithms.” Department of Data Theory FSW/RUL. http://deleeuwpx.net/janspubs/1975/notes/deleeuw_U_75a.pdf.
- De Leeuw, J., and P. Mair. 2009. “Homogeneity Analysis in R: The Package Homals.” *Journal of Statistical Software* 31 (4): 1–21. http://deleeuwpx.net/janspubs/2009/articles/deleeuw_mair_A_09a.pdf.
- Gifi, A. 1990. *Nonlinear Multivariate Analysis*. New York, N.Y.: Wiley.
- Kruskal, J.B., and J.D. Carroll. 1969. “Geometrical Models and Badness of Fit Functions.” In *Multivariate Analysis, Volume II*, edited by P.R. Krishnaiah, 639–71. North Holland Publishing Company.
- Lange, K. 2016. *MM Optimization Algorithms*. SIAM.
- Michailidis, G., and J. De Leeuw. 1998. “The Gifi System for Descriptive Multivariate Analysis.” *Statistical Science* 13: 307–36. http://deleeuwpx.net/janspubs/1998/articles/michailidis_deleeuw_A_98.pdf.
- Rockafellar, R.T. 1970. *Convex Analysis*. Princeton University Press.